

II) Interactions with Sumex-Aim resource

Collaborations and medical use of programs

Dr. Jon Heiser

We have been working with Dr. Jon Heiser of the Department of Psychiatry of the University of California at Irvine, in an effort to create a consultant for the use of psychoactive drugs. We began by creating a version of Mycin that had all of the infectious disease knowledge removed from it, and showed Dr. Heiser how to build up the required base of knowledge about the new field.

He has, with his students, developed a small, but functional system that demonstrates encouraging performance on the task. Work has now begun in earnest to extend the competence of this pilot system, to produce a consultant with a useful level of performance.

It is interesting to note that the explanation capabilities required no modification whatever, and worked in the new system exactly as designed for the original system, despite the change in domains.

INTERNIST Project

The Sumex computer has made possible a valuable interaction between researchers on the MYCIN project at Stanford University and those working on the INTERNIST project at the University of Pittsburgh. These researchers are studying the possible representations and uses for disease models in a medical diagnosis system. Both research groups have been able to run each others programs and to study the medical knowledge bases which are stored on the Sumex computer. Communication between project members has also been greatly facilitated through use of the Sumex system.

Stanford Infectious Disease Faculty

Dr. Victor Yu of our group has been actively soliciting the involvement of the Stanford ID faculty in the development and evaluation of Mycin. He recently presented the system to the faculty and fellows of the Department, and has been seeking ways to involve the system in the Department's educational activities. For instance, medical students under his supervision have used the system during their ID rotation, comparing its results and reasoning process with their own on problems encountered in patients on the wards.

The Pulmonary Function Facility

Members of the Mycin project have also been collaborating with Dr. John Osborn and his co-workers of the Presbyterian Hospital/Pacific Medical Center in San Francisco on the development of a program to interpret the results of standard pulmonary function tests. The program is designed to perform a range of tasks, including: identifying the need to repeat tests because of poor patient effort; identifying the need for additional information in order to make a more definitive diagnosis; reporting and explaining the reasons for primary and secondary diagnoses and severity of any disease state; identifying the relation between diagnosis and any referral diagnosis; and interpreting any change from previous tests, or limitations on the interpretation because of the test methodology and the patient effort.

Sharing with other projects

Groups at Rutgers University, the University of Pittsburgh, Rochester University, and the University of Virginia Medical School have all been involved in varying degrees with running Mycin and evaluating its performance. They have suggested to us improvements in its design, and stock of medical knowledge, and made useful contributions to its development.

In addition, we have made use of the programs developed at both Rutgers and Pittsburgh. The former has been instructive to us in its handling of dynamically changing situations, while the latter has helped us to develop our own ideas about the modelling and use of prototypical descriptions of disease states.

The Molgen group at Stanford has also profited from much of our experience in acquiring knowledge and building large knowledge bases. Several of their

techniques for accumulating knowledge about genetics are based on extensions to ideas first suggested in some of our work.

In all of these cases, the use of Sumex as a national resource has clearly been a critical factor in making possible this sort of interaction.

References

- [1] Reiman H H, D'ambola J, The use and cost of antimicrobials in hospitals, Arch Environ Health, 13:631-636 (1966).
- [2] Kunin C M, et.al., Use of antibiotics: a brief exposition of the problem and some tentative solutions, Anns Int Med, 79:555-560 (1973).
- [3] Sheckler W E, Bennett J V, Antibiotic usage in seven community hospitals, J Amer Med Assoc, 213:264-267 (1970).
- [4] Roberts A W, Visconti J A, The rational and irrational use of systemic antimicrobial drugs, Amer J Hosp Pharm, 29:828-834 (1972).
- [5] Simmons H E, Stolley P D, This is medical progress? Trends and consequences of antibiotic use in the United States, J Amer Med Assoc, 227:1023-1026 (1974).
- [6] Kagan B M, Fanin S L, Bardie F, Spotlight on antimicrobial agents, JAMA, 226:306-310 (1973).

6.1.5 PROTEIN STRUCTURE PROJECT

Protein Structure Modeling Project

Prof. J. Kraut and Dr. S. Freer (Chemistry, U. C. San Diego) and
Prof. E. Feigenbaum and Dr. R. Engelman (Computer Science, Stanford)

I. Summary of research program

A. Technical goals

The goals of the protein structure modeling project are to 1) identify critical tasks in protein structure elucidation which may benefit by the application of AI problem-solving techniques, and 2) design and implement programs to perform those tasks. We have identified two principal areas which have both practical and theoretical interest to both protein crystallographers and computer scientists working in AI. The first is the problem of interpreting a three-dimensional electron density map. The second is the problem of determining a plausible structure in the absence of phase information normally inferred from experimental isomorphous replacement data. Current emphasis is on the implementation of a program for interpreting electron density (e.d.) maps.

B. Medical relevance and collaboration

The biomedical relevance of protein crystallography has been well stated in a recent textbook on the subject (Blundell & Johnson, Protein Crystallography, Academic Press, 1976):

"Protein Crystallography is the application of the techniques of X-ray diffraction ... to crystals of one of the most important classes of biological molecules, the proteins. ... It is known that the diverse biological functions of these complex molecules are determined by and are dependent upon their three-dimensional structure and upon the ability of these structures to respond to other molecules by changes in shape. At the present time X-ray analysis of protein crystals forms the only method by which detailed structural information (in terms of the spatial coordinates of the atoms) may be obtained. The results of these analyses have provided firm structural evidence which, together with biochemical and chemical studies, immediately suggests proposals concerning the molecular basis of biological activity."

The project is a collaboration of computer scientists at Stanford University and crystallographers at the University of California at San Diego (under the direction of Prof. Joseph Kraut) and at Oak Ridge National Laboratories (Dr. Carroll Johnson).

C. Progress summary

During the past year we have been designing and implementing a system of programs for interpreting three-dimensional e.d. maps. Progress has been made by attacking the problem from two directions: working upward from the primary data (i.e. the array of e.d. values) to higher level symbolic abstractions, and working downward from the given amino acid sequence and other experimental information to generate candidate structures which can then be confirmed by the abstracted data.

In the "bottom-up" area of research we have developed and implemented programs for analyzing topological features of the skeletonized e.d. map in terms of protein structural elements (e.g., side chains, chain ends, bridges, etc.), for finding local maxima, and, recently for generating a critical point network, i.e. a three-dimensional spanning tree which connects all critical points (peaks, saddle points) found in the map.

In the "top-down" area we have designed and implemented, in INTERLISP, a structure inference program which generates structural hypotheses at several levels of detail. At present the program can infer, from the amino acid sequence and other chemical information, and the symbolic abstractions of the e.d. map, the location of heavy atoms, cofactors and chain ends. Those features provide footholds, i.e. islands of certainty, from which additional structure is inferred by extension. Work is currently in progress on identification of the main chain, disambiguation of multiply connected regions and classification of side chain regions.

The system under development is knowledge-based. Both the corpus of knowledge of the task domain and the problem-solving strategy knowledge are incorporated as production-like rules.

D. List of Publications

- 1) Robert S. Engelman and H. Penny Nii, "A Knowledge-Based System for the Interpretation of Protein X-Ray Crystallographic Data," Heuristic Programming Project Memo HPP-77-2, January, 1977. (Alternate identification: STAN-CS-77-589)
- 2) E.A. Feigenbaum, R.S. Engelman, C.K. Johnson, "A Correlation Between Crystallographic Computing and Artificial Intelligence," in Acta Crystallographica, A33:13, (1977). (Alternate identification: HPP-77-25)

II. Interaction with the SUMEX-AIM resource

A. Collaborations

The protein structure modeling project has been a collaborative effort since its inception, involving co-workers at Stanford and UCSD (and, more recently, at Oak Ridge). The SUMEX facility has provided a focus for the communication of knowledge, programs and data. Without the special facilities provided by SUMEX the research would be seriously impeded. Computer networking has been especially effective in facilitating the transfer of information. For example, the more traditional computational analyses of the UCSD crystallographic data are made at the CDC 7600 facility at Berkeley. As the processed data, specifically the e.d maps and their Fourier transforms, become available, they are transferred to SUMEX via the FTP facility of the ARPA net, with a minimum of fuss. (Unfortunately, other methods of data transfer are often necessary as well -- see below.) Programs developed at SUMEX, or transferred to SUMEX from other laboratories, are shared directly among the collaborators. Indeed, with some of the programs which have originated at UCSD and elsewhere, our off-campus collaborators frequently find it easier to use the SUMEX versions because of the interactive computing environment and ease of access. Advice, progress reports, new ideas, general information, etc. are communicated via the message and/or bulletin board facilities.

B. Interaction with other SUMEX-AIM projects

Our interactions with other SUMEX-AIM projects have been mostly in the form of personal contacts. We have strong ties to the DENDRAL, Meta-DENDRAL and MOLGEN projects and keep abreast of research in those areas on a regular basis through informal discussions. The SUMEX-AIM workshop in June, 1976 provided an excellent opportunity to survey all the projects in the community. Common research themes, e.g. knowledge-based systems, as well as alternate problem-solving methodologies were particularly valuable to share. (That workshop was very likely the most significant conference for applied AI to be held in 1976.)

6.2 NATIONAL AIM PROJECTS

The following group of projects is formally approved for access to the AIM aliquot of the SUMEX-AIM resource. Their access is based on review by the AIM Advisory Group and approval by the AIM Executive Committee.

6.2.1 ACQUISITION OF COGNITIVE PROCEDURES (ACT)

Acquisition of Cognitive Procedures (ACT)

Dr. John Anderson
Yale University

I. Summary of Research Program

A. Technical goals:

To develop a production system that will serve as an interpreter of the active portion of an associative network. To model a range of cognitive tasks including memory tasks, inferential reasoning, language processing, and problem solving. To develop an induction system capable of acquiring cognitive procedures with a special emphasis on language acquisition.

B. Medical relevance and collaboration:

1. The ACT model is a general model of cognition. It provides a useful model of the development of and performance of the sorts of decision making that occur in medicine.
2. The ACT model also represents basic work in AI. It is in part an attempt to develop a self-organizing intelligent system. As such it is relevant to the goal of development of intelligent artificial aids in medicine.

We have been evolving a collaborative relationship with Dr. James Greeno and Allan Lesgold at the University of Pittsburgh. They are applying ACT to modeling the acquisition of reading and problem solving skills. We plan to make ACT a guest system within SUMEX. ACT is currently at the state where it can be shipped to other INTERLISP facilities. We have received a number of inquiries about the ACT system. ACT is a system in a continual state of development but we periodically freeze versions of ACT which we maintain and make available to the national AI community.

C. Progress and accomplishments:

ACT provides a uniform set of theoretical mechanisms to model such aspects of human cognition as memory, inferential processes, language processing, and problem solving. ACT's knowledge base consists of two components, a propositional component and a procedural component. The propositional component is provided by an associative network encoding a set of facts known about the world. This provides the system's semantic memory. The procedural component

consists of a set of productions which operate on the associative network. ACT's production system is considerably different than many of the other currently available systems (e.g., Newell's PSG). These differences have been introduced in order to create a system that will operate on an associative network and in order to accurately model certain aspects of human cognition.

A small portion of the semantic network is active at any point in time. Productions can only inspect that portion of the network which is active at the particular time. This restriction to the active portion of the network provides a means to focus the ACT system in a large data base of facts. Activation can spread down network paths from active nodes to activate new nodes and links. To prevent activation from growing continuously there is a dampening process which periodically deactivates all but a select few nodes. The condition of a production specifies that certain features be true of the active portion of the network. The action of a production specifies that certain changes be made to the network. Each production can be conceived of as an independent "demon." Its purpose is to see if the network configuration specified in its condition is satisfied in the active portion. If it is, the production will execute and cause changes to memory. In so doing it can allow or disallow other productions which are looking for their conditions to be satisfied. Both the spread of activation and the selection of productions are parallel processes whose rates are controlled by "strengths" of network links and individual productions. An important aspect of this parallelism is that it is possible for multiple productions to be applied in a cycle through the set of productions. Much of the early work on the ACT system was focused on developing computational devices to reflect the operation of parallel, strength-controlled processes and working out the logic for creating functioning systems in such a computational medium.

We have successfully implemented a number of small-scale systems that model various psychological tasks in the domain of memory, language processing, and inferential reasoning. A larger scale effort is underway to model the language processing mechanisms of a young child. This includes implementation of a production system to analyze linguistic input, make inferences, ask and answer questions, etc. Also a great deal of effort is being given to developing learning mechanisms that will acquire and organize the productions for this language processing. This learning program attempts to acquire procedures from examples of the computations desired of the procedures. For instance, the program learns to comprehend and generate sentences by being given sentences and picture representations of the meaning of the sentences (actually hand encodings of the pictures). Although this effort is focused on induction of linguistic procedures, the hope is to develop a general model of induction of cognitive procedures and not to place any language-specificity into the induction procedures.

At the time of this report, we have completed the F version of ACT which is the system with learning capabilities. We are currently testing and tuning the system on a number of linguistic examples. Other projects which are progressing in earlier versions of ACT include use of spreading activation to model semantic disambiguation, modeling of the reading process, and modeling of solutions to word arithmetic problems.

D. Current list of project publications:

- [1] Anderson, J.R. Computer simulation of a language acquisition system: A second report. In D. LaBerge and S.J. Samuels (Eds.). Perception and Comprehension. Hillsdale, N.J.: L. Erlbaum Assoc., 1976.
- [2] Anderson, J.R. Language, Memory, and Thought. Hillsdale, N.J.: L. Erlbaum, Assoc., 1976.
- [3] Anderson, J.R. Induction of augmented transition networks. Cognitive Science, 1977, in press.
- [4] Anderson, J.R. & Kline, P. Design of a production system. Paper to be presented at the Workshop on Pattern-Directed Inference Systems, Hawaii, May 23-27, 1977.
- [5] Anderson, J.R., Kline, P. & Lewis, C. Language processing by production systems. To appear in P. Carpenter and M. Just (Eds.). Cognitive Processes in Comprehension. L. Erlbaum Assoc., 1977.
- [6] Kline, P.J. & Anderson, J.P. The ACTE User's Manual, 1976.

II. Interaction With the SUMEX-AIM Resource

The SUMEX-AIM resource is superbly suited for the needs of our project. We have made the most extensive use of the INTERLISP facilities and the facilities for communication on the ARPANET. We have found the SUMEX personnel extremely helpful both in terms of responding to our immediate emergencies and in providing advice helpful to the long-range progress of the project. Despite the fact that we are on the other side of the continent, we have felt almost no degradation in our ability to do research. We find we can easily list on the terminal a small portion of programs under modification. The willingness of SUMEX mail listing has also meant we can keep relatively up-to-date records of all programs under development.

A unique east coast advantage of working with SUMEX is the low loading of the system during the mornings. We have been able to get a great deal of work done during these hours and try to save our computer-intensive work for these hours.

We have found our one AIM work shop so far (1976) a very useful opportunity to meet with colleagues and exchange ideas.

A particularly striking example of the utility of the SUMEX resource was illustrated in the move from Michigan. In the summer of 1976 Anderson moved to Yale and Greeno to Pittsburgh. There was no loss at all associated with having to transfer programs from one system to another. At Yale we were programming the day after we arrived. The SUMEX link has also permitted continued collaboration with Greeno.

6.2.2 CHEMICAL SYNTHESIS PROJECT (SECS)

SECS - Simulation and Evaluation of Chemical Synthesis

W. Todd Wipke
Department of Chemistry
University of California at Santa Cruz

I. Summary of Research Program

A. Technical Goals.

The long range goal of this project is to develop the logical principles of molecular construction and to use these in developing practical computer programs to assist investigators in designing stereospecific syntheses of complex bio-organic molecules. Our specific goals this past year focused on improvement of the library of chemical transforms, completion of the perception of molecular symmetry and integrating the use of symmetry information throughout SECS including the strategy module. We also wanted to improve the execution speed of SECS, and the speed of graphical interaction over remote communication lines. We planned to simplify the program from the user's viewpoint by including automatic file failsafing, improvement of HELP commands, and non-fatal handling of all errors, as well as production of user's manuals for operation of the program and the writing of chemical transforms. Additionally we intended to initiate applications of SECS to the areas of biosynthesis and metabolism of compounds, as well as phosphorus chemistry. Finally we hoped to improve the strategic constraints and controls that guide SECS in growing a synthesis tree.

B. Medical Relevance and Collaboration.

The development of new drugs and the study of how drug structure is related to biological activity depends upon the chemist's ability to synthesize new molecules as well as his ability to modify existing structures, e.g., incorporating isotopic labels into biomolecular substrates. The Simulation and Evaluation of Chemical Synthesis (SECS) project aims at assisting the chemist in designing stereospecific syntheses of biologically important molecules. The advantages of this computer approach over a manual approaches are manifold: 1) greater speed in designing a synthesis; 2) freedom from bias of past experience and past solutions; 3) thorough consideration of all possible syntheses using a more extensive library of chemical reactions than any individual person can remember; 4) greater capability of the computer to deal with the many structures which result; and 6) capability of computer to see molecules in graph theoretical sense, free from bias of 2-D projection.

SECS was designed to be able to apply any kind of chemical transformation, and because of this generality we see SECS finding application in biogenesis and metabolism (see section II A below). The objective of using SECS in biogenesis is to predict possible biogenetic pathways for a given natural product and also

to predict related compounds which might also co-occur in nature. This can be a great aid in searching for new natural products and in structure elucidation.

The objective of using SECS in metabolism is to predict the plausible metabolites of a given xenobiotic in order that they may be analyzed for possible carcinogenicity. Metabolism research may also find this useful in the identification of metabolites in that it suggests what to look for, and in the identification of possible metabolic pathways connecting a metabolite to a xenobiotic.

C. Progress and Accomplishments.

RESEARCH ENVIRONMENT: At the University of California, Santa Cruz, we have a GT-40 graphics terminal connected to the SUMEX-AIM resource by a 1200 baud leased line and a TI 725 thermal printing teletype connected via TYMNET at 300 baud. UCSC has only a small IBM 370/145 and a PDP-11/45 (limit of 12 K words per user) available, both of which are unsuitable for this research. From July until December our research group had to occupy temporary space during renovation, but is now finally in permanent space in Thimann Laboratories where we have close collaboration with other organic chemists.

CHEMICAL TRANSFORMS: The library of chemical transforms has been reorganized and reevaluated during the past year by Mr. Dolata, a student of Professor D.A. Evans of Cal Tech. New reactions were added and the scope and limitations of others were updated and leading references provided. Additionally, Merck, Sharp, and Dohme Research Laboratories provided revisions of many transforms which a group of 25 synthetic chemists had carefully researched.

SYMMETRY: An efficient algorithm for recognizing molecular symmetry was developed last year. This year that algorithm has been tested against all possible molecular point groups and a few problems which developed were corrected. The algorithm has been documented and initial studies begun on actually determining the point group of a molecule. The symmetry group is now utilized in conjunction with the symmetry of a chemical transform so the transform is applied in all possible unique ways, to generate a non-redundant set of precursors. This symmetry of course takes into account stereochemistry of saturated centers and double bonds. We have surveyed literature syntheses for examples of existing heuristics based on symmetry which can be used for automatically generating high level strategies. This information has never been pulled together before and should make an interesting contribution also to organic synthesis.

STRATEGIC CONTROL: Last year we began developing an implementation of strategic control for SECS, and a simple language for expressing strategies independent of chemical transforms. Since these strategies contain expressions which refer to the molecular structure, it was also necessary to incorporate symmetry here too. For example, if a particular bond is designated as strategic to break, but a transform breaks another bond, the strategy is still satisfied if the two bonds are equivalent by symmetry. This problem becomes more complex when pairs of bonds are specified and when there are logical connectives (AND, OR, XOR, and NOT) involved. This has however been solved. Other changes since last year include a completely new user interface to strategy to allow error

correction and very easy modification of goals. Finally quantitative experiments have been performed to measure the effect of developing a synthesis tree with various types of strategic constraints. The net result of this work is that the user can more easily constrain SECS now to work only in areas which the user decides are worthwhile, consequently fewer precursors are generated which the user would delete.

USER INTERFACES: Users of SECS had difficulty understanding how to copy files into work areas in order to save or restore synthesis trees. Now SECS does all file manipulation, eliminating the problem. Further SECS now automatically failsafes the synthesis tree at key points so that in the event of machine or communication failure the user can automatically restart his analysis from the last key point. Considerable modifications were made to the graphical interface for increasing readability and speed of interaction. Over long slow communication lines (which happens to be the way most SECS users are accessing the program) interactive graphics must be done with care, minimizing the amount and frequency of picture transmission, in order to achieve even tolerable man-machine communication. Lastly, we have implemented appropriate input procedures to eliminate the possibility of a fatal crash from user input errors. According to user reports this was a major problem.

PHOSPHORUS CHEMISTRY: Graphical input and output procedures were developed for entering the stereochemical configuration of a trigonal bipyramid (TBP) phosphorus atom and for producing a correct structural diagram from the machine's internal representation. The SEMA algorithm for generating a stereochemically unique name was extended to deal with the 20 possible configurations for each TBP center, including the ability to recognize enantiomers. The ALCHEM language for representing chemical transforms was extended to facilitate manipulation of TBP's, including changes from trigonal and tetrahedral configurations to square base pyramid and TBP. Queries may deal with apicophilicity, and axial or equatorial orientation. The fine details of phosphorus chemistry such as the fact that groups entering or leaving the phosphorus coordination sphere normally do so from the apical position. Pseudo rotation, apicophilicity, and strain energy are considered in evaluating the stable TBP configurations and in checking for identical structures. A library of phosphorus chemistry is now being prepared in collaboration with a group at the University of Strasbourg, France.

COMPUTER-AIDED ELUCIDATION OF BIOGENETIC PATHWAYS: Although a great amount of effort has been spent on various areas of biogenesis, there have been few attempts to develop general techniques for the elucidation of biogenetic schemes. As a result, the formulation of biogenetic schemes has often been criticized for its lack of rigor and explicit criteria. Our approach is to develop general techniques which lead to the postulation of plausible biogenetic pathways, using the SECS as an aide in obtaining and analyzing solutions to this complex problem. It is our hope this application of computer problem solving techniques will not only uncover new ways of recognizing and evaluating biogenetic pathways but also provide added support to deductions made from biogenetic schemes, such as the generality of a scheme which may be tested in only a few species.

With the proper input information and goals well defined there may be explicit rules to guide the chemist to plausible biogenetic pathways for a particular natural product. Unfortunately, the vast majority of solutions to this problem are determined by a combination of the experienced natural products

chemist's ability to consider the most important rules involved and his unique set of experience-based prejudices. There may be some means to represent and utilize all of the known relevant rules, data and possibly even experience-based prejudices to arrive at the best plausible pathways. The most precise method for representing, developing and testing such a theory is in the form of a computer program. To implement such a computer program, known rules and constraints must be clearly defined, then those that are applicable can be applied at each step of the analysis toward the desired goal. This will keep the solution pathways logically pure and insure that all alternatives which satisfy the rules and constraints are considered. This guarantee of completeness simply can not be made using hand analysis.

A new reaction library containing biogenetic transformations have been written. After inputting a natural product the program will apply the biogenetic transforms which fit the natural product. This generates a set of plausible biogenetic precursors to the target natural product. By continuing this process with the precursors generated, the plausible biogenetic pathways for the natural product can be discovered.

The structures of marine natural products were entered into the program and the plausible biogenetic pathways for these compounds were generated and analysed. Biogenetic pathways which had been proposed in the literature were among the pathways discovered, as were other plausible pathways which would now have to be considered. The success we attained in this research effort verified the applicability of the SECS program as an aid in the analysis of metabolic pathways.

COMPUTER-AIDED PREDICTION OF METABOLITES FOR CARCINOGENICITY STUDIES: We have initiated a research project in collaboration with the Chemical Carcinogenesis group at the National Cancer Institute. The objective of this research is to establish a computer program by which a biochemist or metabolism expert can explore the metabolism of a chemical compound. The investigator enters the substrate molecule by interacting with an input and structure editing module.

Then the program will apply the biological transforms which "fit" the structure, taking into consideration all the context information (2-D, 3-D, and electronic) available about the transform and all perceived information about the structure. This will generate a set of metabolites which are one step away from the substrate structure. The metabolites will be ranked according to expected probability or yield. The exact parameters which should be monitored will be determined during the course of this research. An evaluation module may then screen these metabolites according to criteria specified by the investigator. Duplicate metabolites arising from different pathways will be labelled to indicate that fact. Finally the investigator will be shown the set of metabolites together with data about the transform which produced each one and the values of the parameters being monitored.

The investigator may select one metabolite for further metabolism or may request that all be processed for a specified number of steps. In this way a "tree" of metabolites is produced and displayed. The entire state of the user's tree may be saved to permit continuation of the analysis at another time. Exploration of the metabolism tree will be predominately guided interactively by

the expert investigator. We feel that at this stage of development of the field of metabolism and carcinogenicity that interactive guidance by the expert is necessary. There are many areas where the theory is very thin and a given biological transformation may have been observed for only a few substrates. When this transform is applied to a new substrate, some unrealistic metabolites may be generated owing to the deficiency of contextual information and constraints. An expert is necessary to prune the tree and prevent the automatic processing of those unreasonable intermediates. It is much more efficient for the expert to do this pruning as the tree is being grown, rather than later after an enormous tree has been completed.

At some point either during tree generation or at the end, the metabolites will be passed to another program which will identify those metabolites which are identical or "similar" to known carcinogens. Those will be so marked in the tree.

Presently, the major task is the acquisition of the metabolism knowledge base, i.e. the writing of the transformation library to be utilized. Metabolism experts at the National Institute of Health are gleaning this information from both their own research and the metabolism literature. This information will be encoded and the first testing of this new application for the SECS program will begin in June 1977.

D. Current List of Project Publications

- W.T. Wipke and P. Gund, "Simulation and Evaluation of Chemical Synthesis. Congestion: A Conformation Dependent Function of Steric Environment at a Reaction Center. Application with Torsional Terms to Stereoselectivity of Nucleophilic Additions to Ketones," J. Am. Chem. Soc., 98, 8107(1976).
- W.T. Wipke, G. Smith and H. Braun, "SECS-Simulation and Evaluation of Chemical Syntheses: Strategy and Planning," ACS Symposium Proceedings, 1977.
- W.T. Wipke, Computer Planning of Research in Organic Chemistry, Proceedings of the Third International Symposium on Computers in Chemical Education, Research, and Technology, Caracas, Venezuela, 1976.

S.A. Godleski, P.v.R Schleyer, E. Osawa, and W.T. Wipke, "The Systematic Prediction of the Most Stable Neutral Hydrocarbon Isomer," J. Am. Chem. Soc., 99, 0000(1977).

F. Choplin, R. Marc, G. Kaufmann, and W.T. Wipke, "Computer Design of Synthesis in Phosphorus Chemistry. Automatic Treatment of Stereochemistry," J. Am. Chem. Soc., 99, 0000(1977).

Manuals:

SECS Users Manual, June 1976.
SECS Users Guide, Aug 24, 1976.
ALCHEM Tutorial, Sep 21, 1976.

II. Interactions with SUMEX-AIM Resource

A. Examples of Collaborations and Medical use of Programs via SUMEX.

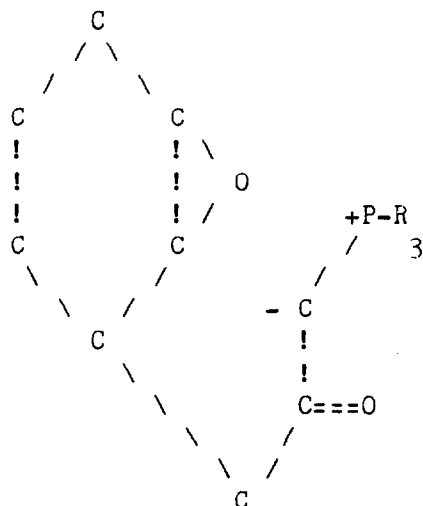
SECS is available in the GUEST area of SUMEX and has been accessed experimentally by many others as well. Professor R. V. Stevens (UCLA) explored some syntheses of lycapodine while visiting Santa Cruz and as a result has requested UCLA to obtain a graphics terminal so he and others at UCLA can access SECS via SUMEX. Professor W. G. Dauben's group (Berkeley) has utilized the SECS model builder on SUMEX is now extending the capabilities of that module of SECS. Mr. Mel Spann of the National Library of Medicine toxicology program is collaborating with us in developing a metabolism library for the metabolism of catechol amines. Also collaborating with us on metabolism are Drs. Ted Gram from Guarino's lab, Harry Gelboin, Dhiren Thakken and Harukiko Hagi from Jerina's lab, Lance Pohl from Gillette's lab, Sidney Nelson from Mitchell's lab, Lionel Poirier from Weisburger's lab, and Ken Chu and Sidney Siegel all of whom are from the National Cancer Institute. Dr. Steve Heller of the EPA and Dr. G.A. Milne of the National Heart and Lung Institute have expressed interest in putting SECS on the Cyphernetics network as a part of the NIH chemical information system. Restrictions on the allowed core image on that system have so far held up the negotiations.

For the past two years SECS has been available over TELENET from First Data Corporation and has been accessed by industry: Squibb; Merck, Sharp and Dohme; Pfizer; Searle; Lederle Labs; FMC; and recently 3M Corporation and Stauffer. Dr. Beryl Dominy of Fizer recently presented a paper before the Pharmaceutical Manufacturer's Association entitled "SECS and the Information Scientist" in which he describes his experiences with SECS, including an example where a synthetic chemist was having difficulty with a particular synthesis, he then went to SECS for possible solutions. SECS suggested another route as being better and indeed that is what he found when he tried it later in the lab.

The availability of SECS on SUMEX-AIM has also served health-related research at the University of California, Santa Cruz. Model building using the SECS model builder is being performed for Professor Edward Dratz (UCSC) to generate conformations of fatty acids isolated from visual membranes ("Structure

and Function of Visual photoreceptors," EI00175), and for Professor Howard Wang (UCSC) to study how conformations of steroids may affect the local anesthetic - membrane interaction ("Role of Membrane Proteins in Local Anesthetic Action," GM22242).

We have assisted Professor J. E. McMurry in his synthetic work towards Aphidicholine and Digitoxigenin by using the model builder for predicting possible reaction pathways. An example is given below, where the conformation of the epoxy-ylide was calculated along with the strain energies of the two possible closure products.



Utilizing the SECS model builder, we have shown that attack on the epoxide to form the fused system should be much more favorable than attack to form the bicyclo compound. Similar studies have been undertaken to predict the stereochemistry resulting from the acid catalyzed cyclization of McMurry's Digitoxigenin precursor (HL-18118 "Total Synthesis of Cardiac Aglycones."): application of SECS using a special library of cationic sigmatropic rearrangement transforms generated the possible products which facilitated identification of some of the side products in the early cyclization experiments. We have also collaborated in the biogenesis work with Professor Phil Crews (UCSC) in marine natural product biogenesis. Dr. Wipke has also used several SUMEX programs such as CONGEN in his course on Computers and Information Processing in Chemistry.

B. Examples of Sharing, Contacts and Cross-fertilization with other SUMEX-AIM projects.

In collaboration with Dr. Ray Carhart and Dr. Dennis Smith of the DENDRAL/CONGEN Project, a Computers in Chemistry Workshop was held at U.C. Santa Cruz on the weekend prior to the Fall 1976 American Chemical Society National Meeting held in San Francisco. The workshop attracted participants representing all parts of the chemical community, academia, industry and government. Morning lecture/discussion sessions introduced the SECS and CONGEN programs running on

SUMEX and the afternoon and evening sessions allowed "hands-on" experience for the participants. The response of the workshop participants was a very positive one with many participants showing so much interest that future collaboration and/or use of the powerful non-numerical computing tools available on SUMEX was discussed.

The SECS project has held joint research group meetings at Stanford with the DENDRAL and AI groups to discuss common problems and research goals. This has been very rewarding since the groups are complementary in orientation. These joint meetings also let the members meet in person after having met on-line on the network. Last year's AIM Conference at Rutgers was also a valuable experience, which allowed us to meet people interested in similar problems in different disciplines. It was particularly useful to have the opportunity to talk with experts designing new languages for knowledge representation and to hear them compare their systems.